Ultra-Reliable Low Latency based on Retransmission and Spatial Diversity in slowly fading channels with co-channel interference

This workshop is being held in conjunction with ETFA 2019.

Ramiro Robles

CISTER-TR-190704

2019/09/10
Ultra-Reliable Low Latency based on Retransmission and Spatial Diversity in slowly fading channels with co-channel interference

Ramiro Robles
CISTER Research Centre
Polytechnic Institute of Porto (ISEP P.Porto)
Rua Dr. António Bernardino de Almeida, 431
4200-072 Porto
Portugal
Tel.: +351.22.8340509, Fax: +351.22.8321159
E-mail: rasro@isep.ipp.pt
https://www.cister-labs.pt

Abstract

This paper presents the analysis of the statistics of latency and information theoretic capacity of an adaptive link with retransmission-spatial diversity in a scenario with co-channel interference. The objective is to evaluate the ability of temporal and spatial diversity to achieve ultra-low values of latency as desired in future 5G and M2M networks with real-time requirements. It is assumed that the source transmits information towards the destination in a Rayleigh fading spatially correlated channel. In case the instantaneous signal-to-interference-plus-noise (SINR) ratio has not surpassed a predetermined reception threshold, then the source engages in a persistent retransmission protocol. All the copies of the original transmission and subsequent retransmissions are stored in memory and processed at the destination using maximum ratio combining (MRC) to obtain a more reliable copy of the signal (a scheme also called retransmission diversity). The retransmission scheme stops once the instantaneous post-processing SINR achieves the desired target threshold. This persistent retransmission scheme can also be regarded as a security mechanism against interference-jamming attacks. Since retransmissions are assumed to take place in a short time interval in order to achieve very low values of latency, contiguous retransmissions are assumed to experience statistical temporal correlation, which is explicitly introduced in the embedded Gaussian channel distribution model. Results suggest that retransmission diversity can provide good latency results in moderate to high values of SINR. However, at low SINR, a combination with other diversity sources will be necessary to achieve the desired target value.
Ultra-Reliable Low Latency based on Retransmission and Spatial Diversity in Slowly Fading Channels with Co-channel Interference

Ramiro Sámano-Robles
Research Centre in Real-time and Embedded Computing Systems, Porto, Portugal
emails:rasro@isep.ipp.pt

Abstract—This paper presents the analysis of the statistics of latency and information theoretic capacity of an adaptive link with retransmission-spatial diversity in an scenario with co-channel interference. The objective is to evaluate the ability of temporal and spatial diversity to achieve ultra-low values of latency as desired in future 5G and M2M networks with real-time requirements. It is assumed that the source transmits information towards the destination in a Rayleigh fading spatially correlated channel. In case the instantaneous signal-to-interference-plus-noise (SINR) ratio has not surpassed a predetermined reception threshold, then the source engages in a persistent retransmission protocol. All the copies of the original transmission and subsequent retransmissions are stored in memory and processed at the destination using maximum ratio combining (MRC) to obtain a more reliable copy of the signal (a scheme also called retransmission diversity). The retransmission scheme stops once the instantaneous post-processing SINR achieves the desired target threshold. This persistent retransmission scheme can also be regarded as a security mechanism against interference jamming attacks. Since retransmissions are assumed to take place in a short time interval in order to achieve very low values of latency, they are modelled with statistical temporal correlation, which is explicitly introduced in the embedded Gaussian channel distribution model. Results suggest that retransmission diversity can provide good latency results in moderate to high values of SINR. However, at low SINR, a combination with other diversity sources will be necessary to achieve the desired target value.

I. INTRODUCTION

Wireless networks are quickly evolving towards the new requirements imposed by the Internet-of-Things, machine-to-machine (M2M) applications, and industry 4.0. Ultra-low latency and high reliability are perhaps the main aspects that new standards like 5G need to achieve in order to support new industrial grade, real time and safety critical services [1]. Since wireless fading, shadowing and interference affect wireless links, successful packet transmissions cannot be hundred percent guaranteed in a deterministic fashion. Retransmissions of the lost (backlog) information must therefore implemented. Since low latency is desired, the less retransmissions are requested to reach the target goal the better latency results will be obtained. Therefore, retransmissions need to be exploited in the best possible way and in the shortest amount of transmission time to minimize the effects over real-time traffic with quasi deterministic deadlines.

This paper explores the ability of wireless links with Rayleigh fading distribution to achieve low latency under a persistent retransmission strategy for packets with low instantaneous channel quality. Since retransmissions are considered to occur over short periods of time, channel outcomes will be considered statistically correlated. All copies of the packet received over contiguous time slots or transmission intervals will be processed using maximum ratio combining (MRC) to obtain a more reliable version of the transmitted information. Co-channel interference is also introduced in the model in an attempt to evaluate the resilience of the retransmission diversity technique in difficult environments (i.e., jamming attacks). A correlated MRC receiver model is used to evaluate the trade-off between spatial and temporal diversity gains.

Retransmission diversity in single user networks has been conventionally analyzed as the well known H-ARQ protocol (Hybrid automatic Repeat Request) [2]. In multi-user settings, retransmission diversity was used for the first time as a method to resolve collisions in random access networks [3]. The protocol was called Network Diversity Multiple Access (NDMA), and it has been analyzed under a multitude of assumptions and combinations over the last two decades [4]-[8]. More recently NDMA has been proposed as a contender of ultra-low latency contention based traffic support in [9].

The main contributions of this paper is the modeling in closed-form of the latency and capacity of an adaptive retransmission diversity link with co-channel interference and spatial-temporal correlation. To the best of our knowledge this is the first approach that obtain this new statistical modelling of the retransmission diversity in the style of a H-AQRQ protocol, but using protocol expressions borrowed from the NDMA literature.

Notation: vector (e.g. x) and matrix variables (e.g. A) are denoted by bold lower and upper case letters, respectively. \( f_{XY}(x) \), \( F_{XY}(x) \), and \( \bar{F}_{XY}(x) \) denote, respectively, the probability density, cumulative distribution, and complementary cumulative distribution function of the random variable \( x \) conditioned on the random variable \( y \). \((\cdot)^H\) and \((\cdot)^T\) are the transpose and Hermitian vector transpose operators.
respectively. \( \omega \) is the frequency domain variable in radians, and \( i = \sqrt{-1} \). \( \Psi_{x|x}(\omega) \) is the characteristic function (CF) of the random variable \( x \) conditioned on a value of the random variable \( y \). \( I_n \) denotes the identity matrix of order \( n \), \( 0_n \) and \( I_n \) denote the column vectors, respectively, of zeros and ones of order \( n \). \( E[\cdot] \) is the statistical average operator.

This paper is organized as follows. Section II describes the system model and the assumptions of the paper. Section V presents analytic results and sketches of the statistics of packet reception. Finally, Section VI presents the conclusions of the paper.

II. SYSTEM MODEL

A. Scenario and signal model

Consider the wireless link depicted in Fig. 1 with one source node, one destination node, and a source of co-channel interference. The destination is assumed to have a multiple antenna receiver with \( M \) elements. The channel between source and the \( m \)th antenna element at the destination node in time slot \( n \) is denoted by \( h_m(n) \) and will be modeled as a circular complex Gaussian random variable with variance \( \gamma: h_m(n) \sim \mathcal{CN}(0, \gamma) \). The indexes \( m \) and \( \tilde{n} \) will be used to enumerate the antennas, and the indexes \( n, \tilde{n} \) and \( \tilde{n}_0 \) enumerate the time slots of a transmission period. The interference term will be considered constant through the transmission period and will also be modeled as a Gaussian circular complex variable with variance \( \sigma_c^2 \). A transmission period is the set of time slots required for a packet transmission to be received correctly at the destination. Since the proposed system uses persistent retransmission to achieve the desired channel quality, a transmission period is composed of a random number of time slots: the original transmission plus the necessary retransmissions.

![Fig. 1. Wireless link with spatial and retransmission diversity with co-channel interference.](image)

The signal transmitted by the source node is denoted by the vector variable \( w \) of \( Q \) symbols: \( w = [w(0), w(1), \ldots, w(Q-1)]^T \). Therefore, the signal received by the \( m \)th antenna at the destination in time slot \( n \) of a transmission period \( \left[ r_m(n)^{(0)}, r_m(n)^{(1)}, \ldots, r_m(n)^{(Q-1)} \right]^T \) can be expressed as follows:

\[
 r_m(n) = h_m(n)w + v_m(n) + \sqrt{\sigma_I}I_Q,
\]

where \( v_m(n) \) is the column vector of zero-mean complex circular Gaussian noise with co-variance matrix \( \Sigma_c^2= v \sim \mathcal{CN}(0_Q, \sigma_c^2I_Q) \), where \( 0_Q \) and \( I_Q \) are, respectively, the vector of zeros and the identity matrix of order \( Q \) and \( \sigma_c^2 \) is the noise variance.

Let us now consider that the transmitted symbol complies with the following power constraint: \( E[w^Hw] = P \), where \( \langle \cdot \rangle^H \) is the Hermitian vector transpose operator. The receiver uses MRC to process the signals received by all the antennas in all time slots. This leads to the well-known formula for the post-processing or instantaneous signal-to-interference-plus-noise ratio (SINR) of the MRC receiver considering perfect channel estimation at the receiver side [10]:

\[
\Gamma_n = \frac{\sum_{\tilde{n}=1}^{n} \sum_{m=1}^{M} P|h_{m}(\tilde{n})|^2}{1 + \sigma_c^2}.
\]

It is assumed that a packet is correctly received by the intended destination is the instantaneous SINR exceeds the reception threshold \( \beta \):

\[
\Pr\{\Gamma_n > \beta\} = \Pr\{\text{packet correctly received by vehicle } i \text{ in time-slot } n\}.
\]

This assumption is mainly intended to activate the retransmission of packets. In practice, packets need to be processed and after hard decision decoding it is possible to evaluate if the packet has errors or not via a redundancy check or other mechanisms. Packets that still contain errors after this stage are considered to be handled by upper layer error control protocols.

For convenience, all channels will be expressed using a linear correlation model similar to the one used in [10], which in our context will be written as follows:

\[
 h_{m}(n) = \sqrt{1 - \lambda_{Z,m,n}^2} Z_{m,n} + \lambda_{m,n} G,
\]

where the variables \( Z_{m,n} \), \( G \) are identically and independently distributed (i.i.d.) zero-mean complex circular symmetrical Gaussian random variables with variance \( \gamma \), and \( \lambda_{m,n} = e^{i \frac{2\pi n}{Nco}} e^{i \frac{2\pi m}{Mco}} \), so that the correlation coefficient can be defined as follows:

\[
\rho_{m,n} = \frac{E[h_{m}(n)^*h_{\tilde{n}}(\tilde{n})]}{\gamma^2} = \lambda_{m,n}^*\lambda_{\tilde{n},\tilde{n}} = e^{i \frac{2\pi (\tilde{n}-n)}{Nco}} e^{i \frac{2\pi (\tilde{n}-m)}{Mco}},
\]

where \( Nco \) and \( Mco \) can be adjusted for different temporal and spatial statistics of the channel correlation. The list of main variables is shown in Table I.
TABLE I
LIST OF MAIN VARIABLES.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>$h_m(n)$</td>
<td>Channel between TX and antenna $m$ in time slot $n$.</td>
</tr>
<tr>
<td>$M$</td>
<td>Number of antennas at the RX.</td>
</tr>
<tr>
<td>$e$</td>
<td>Transmission interval indicator</td>
</tr>
<tr>
<td>$l$</td>
<td>Length of an Tx period</td>
</tr>
<tr>
<td>$\beta$</td>
<td>Reception threshold</td>
</tr>
<tr>
<td>$P$</td>
<td>Tx Power</td>
</tr>
<tr>
<td>$M_{co}$</td>
<td>Spatial correlation parameter</td>
</tr>
<tr>
<td>$N_{co}$</td>
<td>Temporal correlation parameter</td>
</tr>
<tr>
<td>$\rho_{m,n}$</td>
<td>Correlation coefficient between the signal of antenna $m$ with the signal of antenna $n$.</td>
</tr>
<tr>
<td>$\gamma$</td>
<td>Channel variance</td>
</tr>
<tr>
<td>$\sigma_w^2$</td>
<td>Noise variance</td>
</tr>
<tr>
<td>$\sigma_I^2$</td>
<td>Interference variance</td>
</tr>
<tr>
<td>$\Gamma$</td>
<td>Signal-to-Interference-plus-Noise Ratio</td>
</tr>
<tr>
<td>$w$</td>
<td>Signal transmitted by terminal</td>
</tr>
<tr>
<td>$r_m$</td>
<td>Rx Signal in antenna $m$</td>
</tr>
<tr>
<td>$v_m$</td>
<td>Noise vector in antenna $m$</td>
</tr>
</tbody>
</table>

B. Protocol operation

This paper focuses on the evaluation of the statistics of latency of a packet at the head of the queue in the source node to be transmitted correctly (including retransmissions) towards the destination. If the packet initially transmitted by the source node is not correctly received, then it is assumed that an immediate retransmission is requested from the source node. This process is assumed to occur instantaneously and without error. All the copies of the signal transmitted and retransmitted by the source node are stored in the memory at the destination, where they are processed using maximum ratio combining (MRC) to produce a more reliable copy of the original transmission. The process is repeated continuously until the signal is assumed to be correctly received by the destination. While retransmissions are assumed to occur instantaneously in subsequent time slots after the first transmission, all results in this paper can be adapted to consider delay of the feedback channels. The main idea of the instantaneous retransmission diversity scheme proposed in this paper is to evaluate the ability of the wireless component to achieve the lowest possible latency of transmission. In this manner, it will be possible to estimate the performance of the physical layer algorithms to achieve values of latency that can be considered as ultra-low and thus be useful in several types of scenarios with real-time or quasi-real time requirements. We note here that the latency analysis in this paper focuses exclusively on the wireless transmission component. The full analysis of latency end-to-end must include other aspects such as queuing delay, traffic arrival distribution, collision avoidance, etc. The examples in Fig. 1 illustrate different instances of the retransmission scheme. The first transmission period consists of three time slots or transmission time intervals (TTIs), mainly because the channel quality was not good enough in the previous two channel outcomes. The second transmission period has only one time slot, which means the channel quality was good enough and no retransmissions were necessary. The final transmission period consist of two time slots.

III. STATISTICS OF CORRECT SIGNAL RECEPTION

Let us substitute the correlation model described by (4) in the expression of the SINR in (2). This yields:

$$\Gamma_n = \frac{\xi_n}{I + \sigma_v^2},$$

where

$$\xi_n = \sum_{\bar{n}=1}^{M} \sum_{m=1}^{M} P \left| \sqrt{1 - \lambda_m^2} Z_{\bar{m},\bar{n}} + \lambda_{m,n} G \right|^2$$

Consider now the previous expression conditional on an instance of the random variable $G$. Under this assumption, the expression in (6) becomes the summation of the squares of Gaussian complex variables $P \left| 1 - \lambda_m^2 Z_{m,n} \right|^2$ each one with a mean given by $\sqrt{P} \lambda_m G$. Therefore, the variable $\xi_n$ conditional on an instance of random variables $G$ has a non-central chi-square distribution with $n$ degrees of freedom. The characteristic function (CF) of $\xi_n$ conditional on a value of $G$ can be thus written as (see [11] for details of chi-square distributions):

$$\Psi_{\xi_n|G}(i\omega) = \prod_{n=1}^{n} \prod_{m=1}^{M} \frac{1}{(1 - i\omega \gamma_m,n)} e^{i\omega \lambda_m G \gamma_m,n^2},$$

where $\gamma_m,n = P \sum_{\bar{n}=1}^{M} \left( 1 - \lambda_m^2 \right) \gamma_{m,n}$ and $\lambda_m,n = \sqrt{P} \lambda_m G$. Averaging the previous expression over the PDF of the random variable $G$ yields the expression for the unconditioned characteristic function:

$$\Psi_{\xi_n}(i\omega) = \sum_{\bar{n}=1}^{M} \prod_{m,n \neq \bar{n},\bar{n}} \left( 1 - i\omega \gamma_m,n \right)$$

$$\times \prod_{m,n} (1 - i\omega \gamma_m,n) \left( 1 + i\omega \sigma_v^2 \right)^{-1}.$$

Let us now rewrite the reception probability expression in (3) using the explicit expression for the SINR in (5) as follows:

$$\Pr \{ \xi_n > \beta \} = \Pr \left\{ \frac{\xi_n}{I + \sigma_v^2} > \beta \right\} = \Pr \{ \xi_n > \beta(I + \sigma_v^2) \}$$

$$\Pr \{ \xi_n - \beta > \beta \sigma_v^2 \} = \Pr \{ \theta_n > \beta \sigma_v^2 \}$$

Due to the independence assumption of incumbent signals and interference components, the characteristic function of the random variable $\theta$ can be obtained as follows:

$$\Psi_{\theta_n}(i\omega) = \Psi_{\xi_n}(i\omega)(1 + i\omega \sigma_v^2)^{-1}$$

By substituting the CF of the random variable $\xi_n$ we get:

$$\Psi_{\theta_n}(i\omega) = \sum_{\bar{n}=1}^{M} \prod_{m,n \neq \bar{n},\bar{n}} \left( 1 - i\omega \gamma_m,n \right)$$

$$\times \prod_{m,n} (1 - i\omega \gamma_m,n) \left( 1 + i\omega \sigma_v^2 \right)^{-1}.$$
Using partial fraction expansion, the previous expression becomes:

$$
\Psi_{\theta_n}(i\omega) = \sum_{k=1}^{nM+1} \frac{A_k}{i\omega - \gamma_k},
$$

where $A_k = \prod_{k\neq k} (\gamma_k - \gamma_k)^{-1}$, and $-\gamma_k^{-1}$ is the $k$th root of the polynomial function of the denominator. The back-transform of (8) yields a complementary cumulative distribution function (CCDF) given by:

$$
\bar{F}_{\Gamma_n}(y) = \sum_{k=1}^{nM+1} A_k e^{\gamma_k y}.
$$

IV. LATENCY AND CAPACITY STATISTICS

Each packet transmission is considered as correctly received if and only if the instantaneous SINR if above the designed threshold $\beta$. If this is not achieved an immediate retransmission is provided by the source node. Retransmissions are stored in memory and combined to improve SINR conditions. The probability mass function of this number of attempts for a packet to be correctly received can be therefore written as follows:

$$
\Pr\{l = n\} = \prod_{n=1}^{n-1} F_{\Gamma_n|\Gamma_{n-1} < \beta}(\beta) \bar{F}_{\Gamma_n|\Gamma_{n-1} < \beta}(\beta).
$$

By using Bayes theorem it can be proven that:

$$
F_{\Gamma_n|\Gamma_{n-1} < \beta; \eta_n}(\Gamma_n) = \frac{F_{\Gamma_n|\Gamma_{n-1} < \beta}(\beta)}{\Gamma_{n-1}|\eta_n}(\beta).
$$

By substituting the previous expression back in (10) we obtain:

$$
\Pr\{l = n\} = F_{\Gamma_n}(\beta) - F_{\Gamma_n}(\beta).
$$

By using the expression in (7) we obtain:

$$
\Pr\{l_t = n\} = F_{\theta_{n-1}}(\beta^2) - F_{\theta_n}(\beta^2).
$$

Ultra-low values of latency are one of the main attributes of future 5G systems. It is expected that values as low as 1ms will be the standard to achieve ultra low latency. 5G standards are focused on defining transmission frames in the order of 1ms [12]. The capacity can be obtained as the ratio of the average of the resulting SINR over all possible realizations of transmission periods and channel states to the average number of transmission attempts. This can be written as follows:

$$
C_T = \frac{\sum_{k=1}^{\infty} \Pr\{l = k\} C(\Gamma_k, \Gamma_{k-1} < \beta)}{\sum_{k=1}^{\infty} k \Pr\{l = k\}},
$$

where $C(\Gamma_k, \Gamma_{k-1} < \beta) = E[\log_2(\Gamma_k|\Gamma_{k-1} < \beta)]$. The capacity formula reflects the fact that retransmissions of the same information are adaptive and therefore information is not transmitted in a single transmission interval, but in a random number of transmissions and retransmissions.

V. RESULTS

We consider a transmission scenario with different values of Tx power, with constant normalized power ($P = 1$) and two cases with and without interference. The interference variance is set to $\sigma_I^2 = 10$ and noise variance normalized to one. Also different values of the correlation structure in space and time have been considered. The results for latency distribution without and with interference are shown in Fig. 2 and Fig. 3 with a very large detection threshold set to $\beta = 100$. This value was chosen to investigate a very low SNR scenario. The effects of time and spatial correlation are observed in all figures increasing latency. In general, the tail part of the latency distribution is more affected by larger values of $N_{co}$ and $N_{co}$. The results with a detection threshold set to $\beta = 10$ are shown in Fig. 4. Results of capacity versus latency for variable reception threshold $\beta = 1 : 1 : 100$ with and without interference. It can be shown that in the case of interference massive MIMO operation will allow the reduction to almost zero retransmissions, which is desirable in 5G scenarios.

The results with interference confirm the main results found in the scenario without interference, but with increased latency. It can be observed that a good trade-off can be found between the number of antennas and the transmission power to obtain low values of latency. For very high values of antennas as it is expected in massive MIMO systems, the rejection against interference with ultra-low values of latency and high capacity are shown to be relatively easy to be achieved.

![CCDF](image_url)

Fig. 2. CCDF of Latency with zero interference using different numbers of antennas, retransmission diversity and different values of channel correlation parameters with reception threshold $\beta = 100$.

VI. CONCLUSIONS

This paper has presented an analysis of the abilities of spatial-temporal diversity in correlated fading environments to achieve ultra low latency. The main results suggest that in environments with correlated fading and high interference, massive MIMO technology can reduce latency to a minimum number of retransmissions. In other environments, a moderate number of spatial antennas followed by retransmission diversity is enough to keep latency in low values and still achieve high values of capacity. Further analysis of the results presented...
Here considering traffic statistics and multi-user environments is expected as future work.

REFERENCES


